**Alexandria University**
**Faculty of Engineering**
**Computer and Systems Engineering**
**First Year**

**CS 111: Probability Theory and its Applications**
**Fall 2021**

# Lab 4

## Overview

In this lab, you will discover Binomial and Poisson distributions and you'll be exposed to R statistical capabilities by studying various built in functions designed specially for statistical distributions.

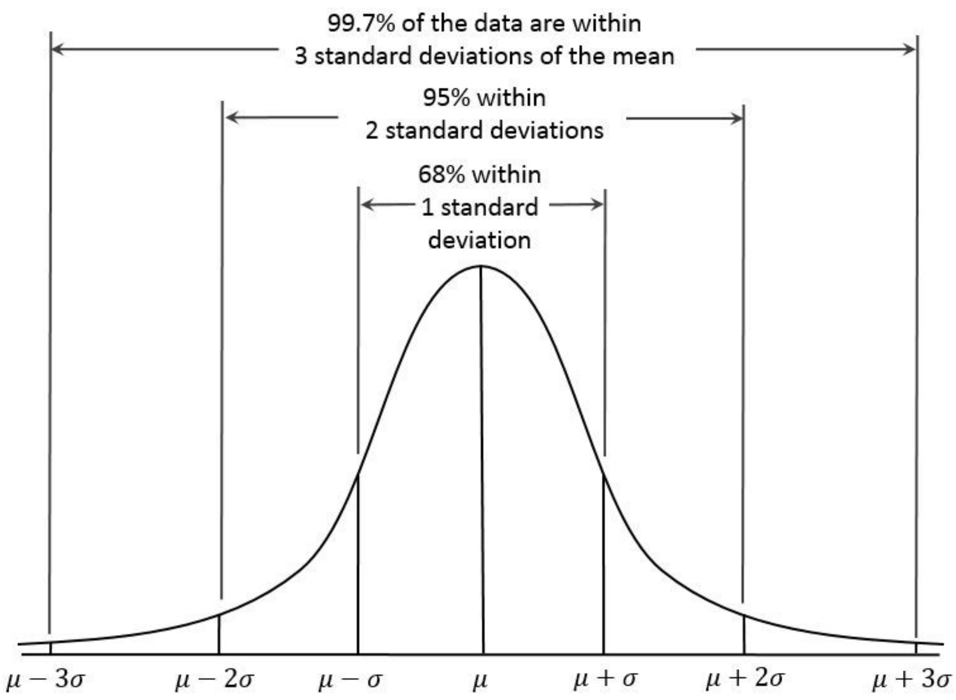
## Statistical Introduction

### Basic Notations:

- **Mean** of a vector is the sum of values over the vector size. It is denoted by $\mu$ or **μ**
- **Median** is the number in the middle of a vector of values ( take care: even/odd cases)
- **Mode** is the value that has the maximum number of occurrences.
- **Variance** measures how far a set of (random) numbers are spread out from their average. It is Often represented by **Var(x)** or $^2$ or $s^2$
- **Standard deviation** is a measure of the amount of variation or dispersion of a set of values. The standard deviation of a random variable, statistical population, data set, or probability distribution is the square root of its variance and is denoted by or *s*

### Normal distribution:

Normal distributions are important in statistics and are often used in the natural and social sciences to represent real-valued random variables whose distributions are not known. In a random collection of data from independent sources, it is generally observed that the distribution of data is normal. Which means, on plotting a graph with the value of the variable in the horizontal axis and the count of the values in the vertical axis we get a bell shaped curve. The center of the curve represents the mean of the data set. In the graph, fifty percent of values lie to the left of the mean and the other fifty percent lie to the right of the graph. This is referred as normal distribution in statistics.




R has four in built functions to generate normal distribution. They are described below.

```
dnorm(x, mean, sd)
pnorm(x, mean, sd)
qnorm(p, mean, sd)
rnorm(n, mean, sd)
```

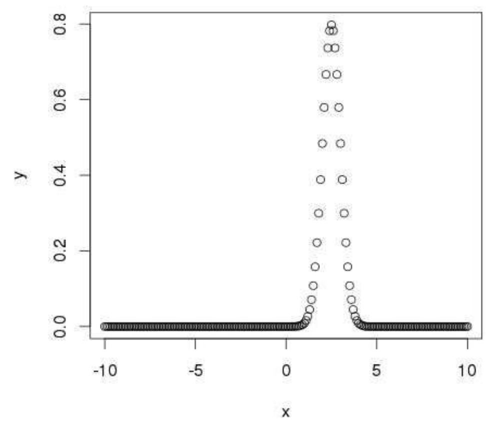
Following is the description of the parameters used in the above functions:

- x is a vector of numbers.
- p is a vector of probabilities.
- n is number of observations(sample size).
- mean is the mean value of the sample data. It's default value is zero.
- SD is the standard deviation. It's default value is 1.

## 1. dnorm()

This function gives the height of the probability distribution at each point for a given mean and standard deviation.

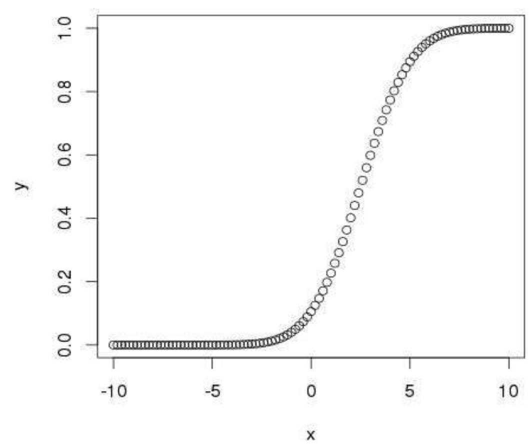
```
# Create a sequence of numbers
between -10 and 10 incrementing by
0.1.
x <- seq(-10, 10, by = .1)
# Choose the mean as 2.5
and standard deviation as
0.5.
y <- dnorm(x, mean = 2.5, sd =
0.5)
# Give the chart file a name.
png(file = "dnorm.png")
plot(x,y)
# Save the file.
dev.off()
```




## 2. pnorm()

This function gives the probability of a normally distributed random number to be less than the value of a given number. It is also called "Cumulative Distribution Function"

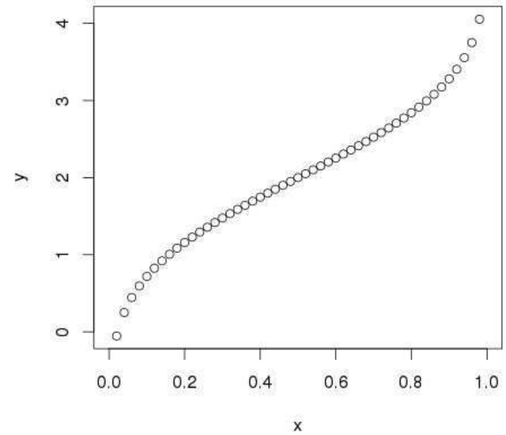
```
# Create a sequence of numbers
between
 -10 and 10 incrementing by 0.2.
x <- seq(-10,10,by = .2)
# Choose the mean as 2.5 and
standard
deviation as 2.
y <- pnorm(x, mean = 2.5, sd = 2)
# Give the chart file a name.
png(file = "pnorm.png")
# Plot the graph.
plot(x,y)
# Save the file.
dev.off()
```



## 3. qnorm()

This function takes the probability value and gives a number whose cumulative value matches the probability value.

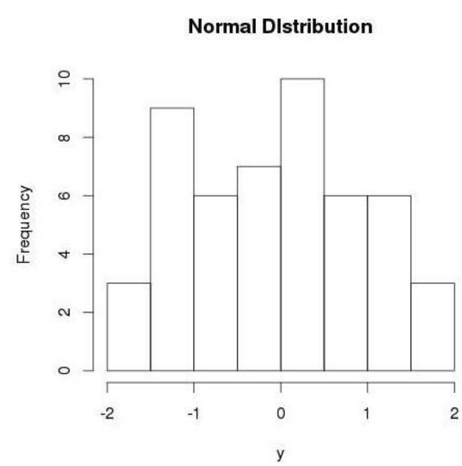
```
# Create a sequence of
probability
values incrementing by 0.02.
x <- seq(0, 1, by = 0.02)
# Choose the mean as 2 and
standard deviation as 3.
y <- qnorm(x, mean = 2, sd = 1)
# Give the chart file a name.
png(file = "qnorm.png")
# Plot the graph.
plot(x,y)
# Save the file.
dev.off()
```




## 4. rnorm()

This function is used to generate random numbers whose distribution is normal. It takes the sample size as input and generates that many random numbers. A histogram is drawn to show the distribution of the generated numbers.

```
# Create a sample of 50 numbers
 which are normally distributed.
y <- rnorm(50)
# Give the chart file a name.
png(file = "rnorm.png")
# Plot the histogram for this
sample.
hist(y, main = "Normal
DIstribution")
# Save the file.
dev.off()
```



***Hints:***

*`png(file = "name.png")` is used to rename the distribution plot*

*`plot(x,y)` is used to plot the distribution*

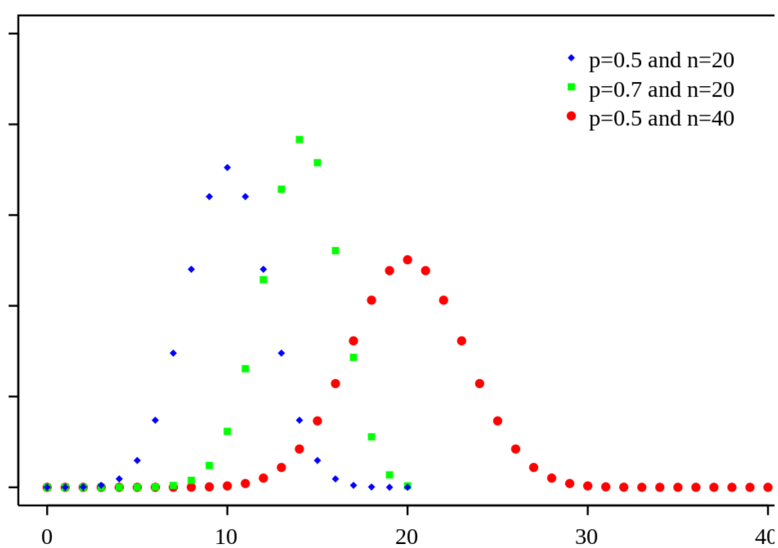
*`dev.off()` is used to save the file*

*`nrow()` counts the number of rows*

# Binomial distribution:

The **binomial distribution** model deals with finding the probability of success of an event which has only two possible outcomes in a series of experiments. For example, tossing of a coin always gives a head or a tail. The probability of finding exactly 3 heads in tossing a coin repeatedly for 10 times is estimated during the binomial distribution.

$$f(x) = \binom{n}{x} p^x (1-p)^{(n-x)} \quad where\ x = 0, 1, 2, ..., n$$



R has four built in functions to represent poisson distribution functions. They are described below.

```
dbinom(x, size, prob)
pbinom(x, size, prob)
qbinom(p, size, prob)
rbinom(n, size, prob)
```

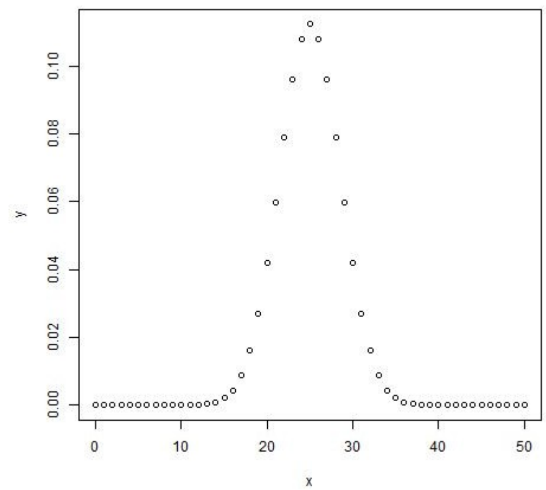
Following is the description of the parameters used in the above functions:

- x is a vector of numbers.
- p is a vector of probabilities.
- n is number of observations.
- size is the number of trials.
- prob is the probability of success of each trial.

## 1. dbinom()

This function gives the probability density distribution at each point.

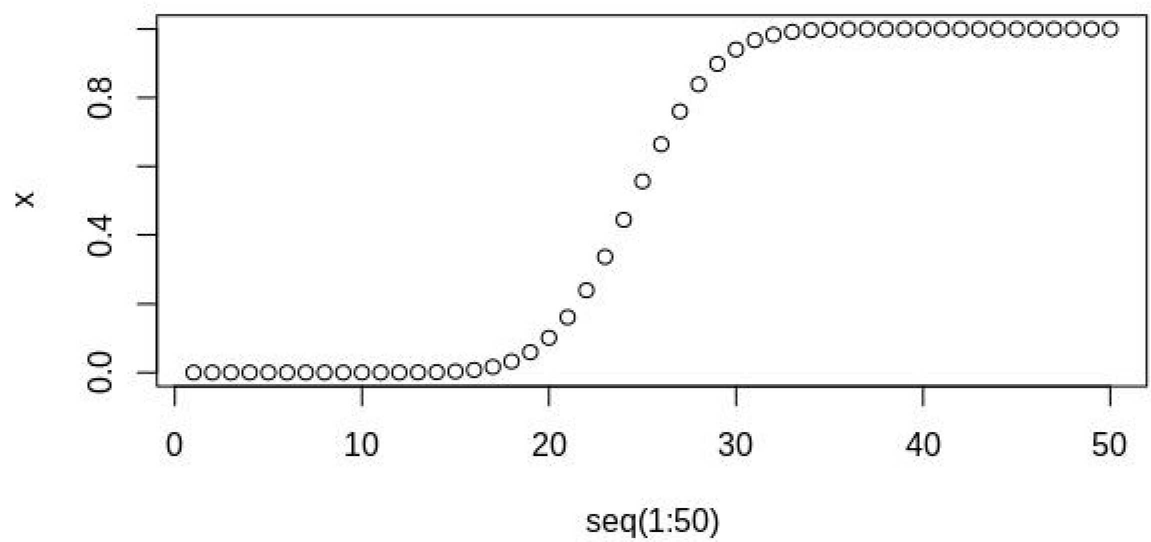
```r
# Create a sample of 50 numbers which are
incremented by 1.
x <- seq(0,50,by = 1)
# Create the binomial distribution.
y <- dbinom(x,50,0.5)
# Plot the graph for this sample.
plot(x,y)
```



### 2. pbinom()

This function gives the cumulative probability of an event. It is a single value representing the probability.

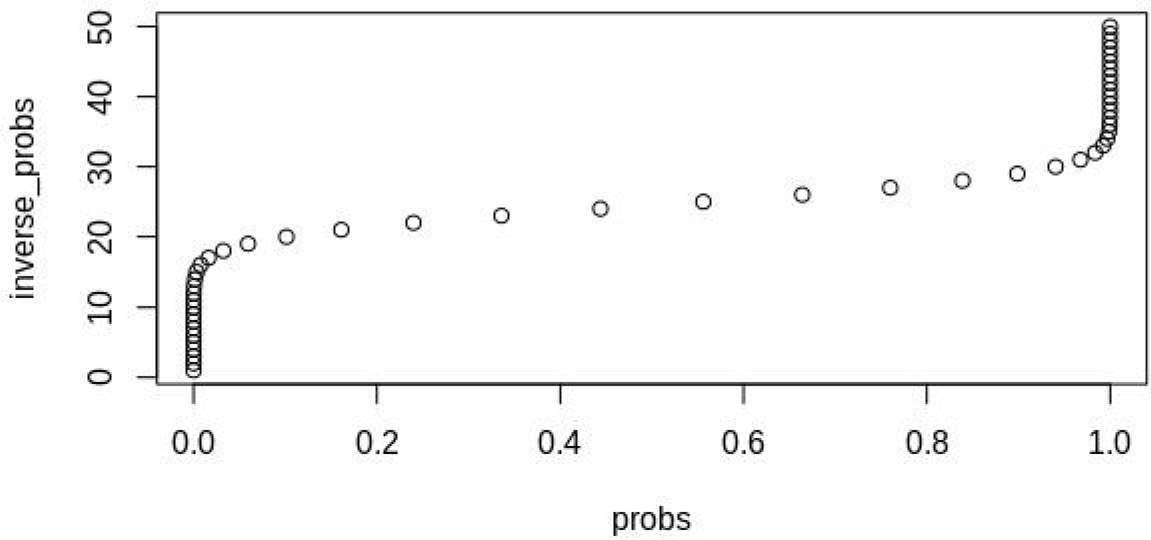
```
# Probability of getting each occurance from 1 : 50 or less in 50 fair
coin tosses
probs <- pbinom(seq(1:50), 50, 0.5)
plot(seq(1:50), probs)
```




### 3. qbinom()

This function takes the probability value and gives a number whose cumulative value matches the probability value.

```
probs <- pbinom(seq(1:50), 50, 0.5)
inverse_probs <- qbinom(x,50,0.5)
plot(probs, inverse_probs)
```



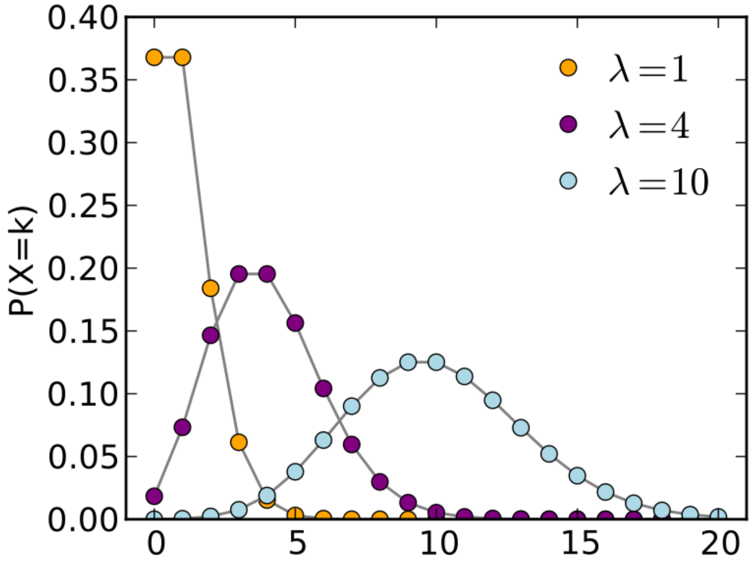

### 4. rbinom()

This function generates required number of random values of given probability from a given sample.

## Poisson distribution:

In probability theory and statistics, the **Poisson distribution** is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time or space if these events occur with a known constant rate and independently of the time since the last event.

$$p(x) = \lambda^x \exp(-\lambda)/x!$$




R has four built in functions to represent poisson distribution functions. They are described below.

```
dpois(q, lambda)
ppois(q, lambda)
qpois(p, lambda)
rpois(n, lambda)
```

Following is the description of the parameters used in the above functions:

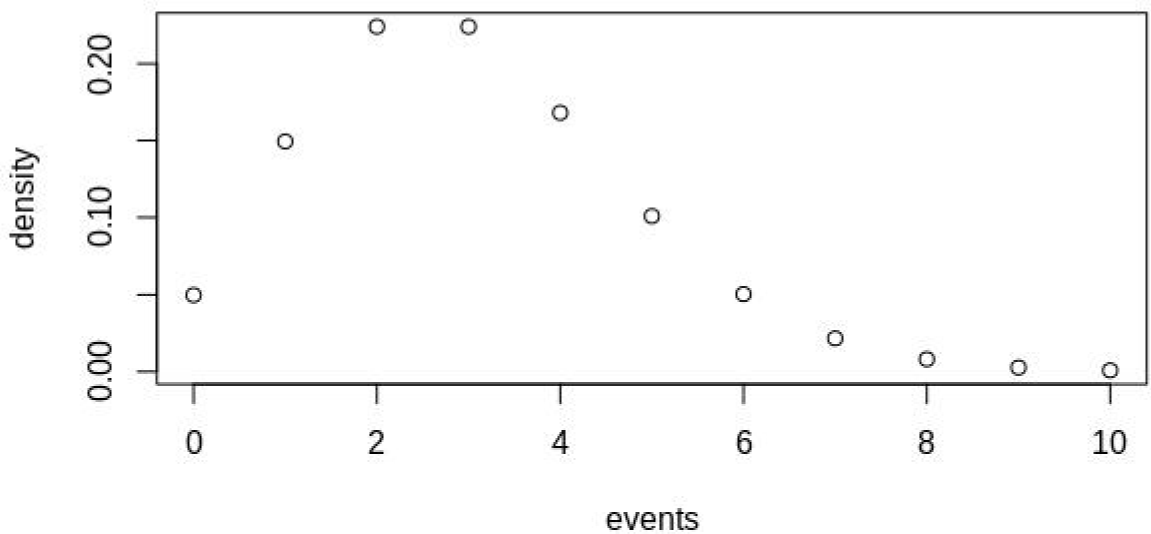
- q is a vector of (non-negative integer) quantiles.
- p is a vector of probabilities.

- n is number of random values to return.
- lambda vector of (non-negative) means..

## 1. dpois()

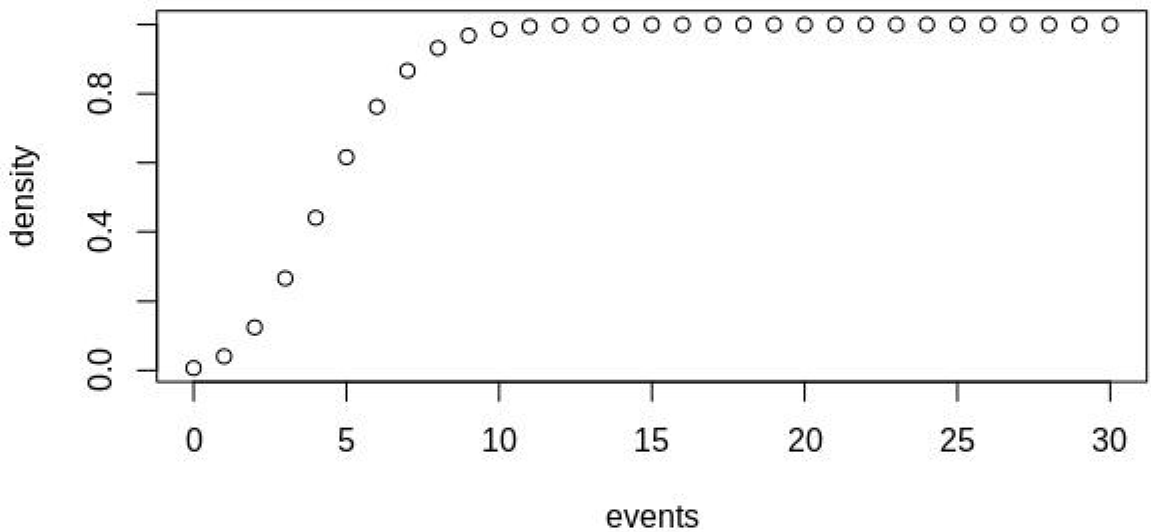
This function gives the height of the probability distribution at each point for a given lambda, ***i.e. Probability Mass Function (PMF)*** values for the given points.

```r
events <- 0:10
density <- dpois(events, 3)
plot(events, density)
```



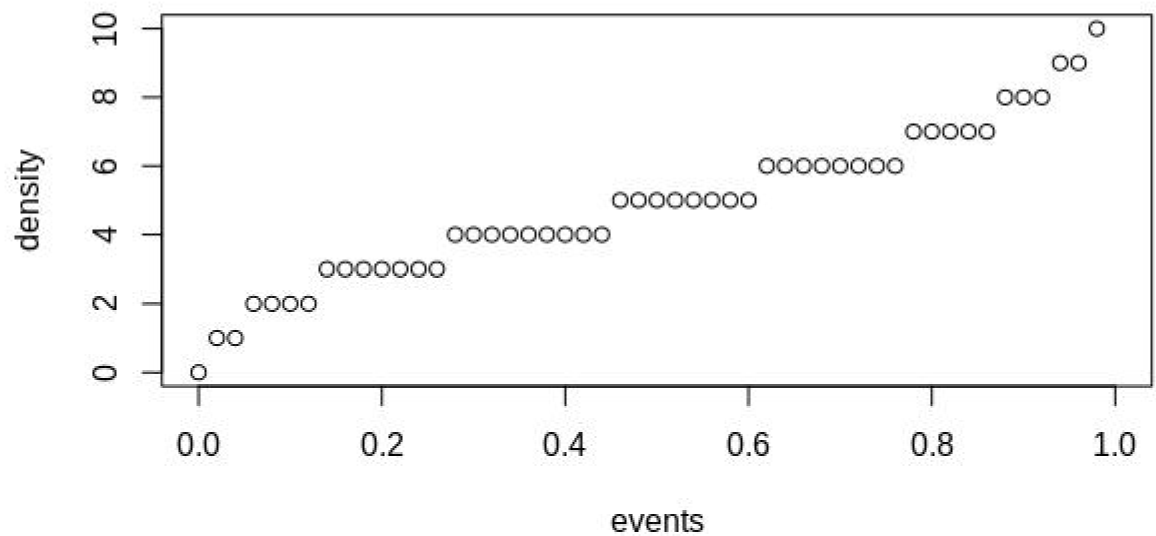

## 2. ppois()

This function gives the probability of a Poisson distributed random number to be less than the values given as ***q*** for a given *lambda*, ***i.e. Cumulative Density Function (CDF)***.




### 3. qpois()

This function takes the probability values vector and gives a number whose cumulative value matches the probability value, ***i.e. CDF Inverse.***




4. **rpois()**

This function is used to generate random numbers whose distribution is Poisson.

# Application

1- Given that the test scores for CSED students fits a normal distribution. Furthermore, the mean test score is 80, with a standard deviation 13.8. Find the percentage of students scoring 75 or more in this exam.

2- What is the probability of getting 26 or less heads from a 51 tosses of a coin?

**3-** How many heads will have a probability of 0.25 will come out when a coin is tossed 100 times? **<u>Note:</u>** Assume that the coin is NOT a fair coin, the probability of giving a head -i.e. success- is **0.6**

4- Assume that you are collecting data for a car selling shop that has sales rate equal to 3 cars per week, What is the probability of making 2 to 4 sales in a week? <u>Use Poisson distribution built-in functions in R.</u>
***Follow up:*** can you solve this problem in 2 different ways? think of the first method as using the PMF and the other method is to use the CDF !